Dose, Resolution and Exposure in Serial Crystallography

Dmitri Starodub, Peter Rez, John Spence and co-workers[1]
Department of Physics, Arizona State University, P.O. Box 871504,Tempe AZ 85287-1504

Resolution of biological diffractive imaging is severely limited by the sample damaging prior to the accumulation of statistically accurate data. In serial crystallography hydrated macromolecules pass successively across an X-ray or electron beam giving a diffraction pattern. The macromolecules are brought into alignment by the electric field of a high power laser. Since the molecule is exposed to the beam for a period of only a few hundred nanoseconds, the probability of an ionization event leading to damage is small, while the scattering from many molecules can be summed to achieve the desired signal. In order to provide a critical dose during molecule transit through an x-ray beam, its intensity should be much above the capabilities of any existing or projected x-ray sources. The effect is the same as having an aligned molecule in the beam for any given time without damage, and the resolution limit is determined only by the exposure time. The purpose of this work is to establish the relationship between exposure time, the required resolution, and the number N of molecules in the beam at any instant.

The real space charge distribution is found using the Hybrid Input Output (HIO) algorithm, where the object is embedded in a support that provides enough constraints to solve the phase problem. To optimize the HIO algorithm the pixel size on the detector should be matched to $2^{1/3}\times D$, where D is the object size, by adjusting the detector position. Then the resolution d in the reconstruction is related to the maximum scattering wave vector, which is governed by the detector size. This means that for a given detector the resolution is fixed both by the geometry and the number of pixels. The counting time can be estimated by setting the minimum acceptable number of counts at the maximum scattering angle, corresponding to the highest spatial frequency or best resolution. An analytic solution for the required exposure can be derived for a spherical macromolecule, and this gives a d^{-4} dependence on desired resolution and an inverse square dependence on wavelength. We estimate that the counting time for 0.5 nm resolution with the 8-kV X-ray flux of 3×10^{8} photons $\mathrm{nm}^{-2}\,\mathrm{s}^{-1}$ (for the projected Energy Recovery Linac source) is 470 s for $N = 1$, which is sufficient to delineate secondary structure.

This approach gives a lower limit for the required X-ray fluence, because it does not account for convergence stability of a phase retrieval algorithm with respect to a shot noise of the scattered photons. For more realistic estimations, we perform simulations of the diffraction pattern for GroEL-GroES, a chaperone protein complex, using coordinates taken from Protein Data Bank (1SVT). After adding a shot noise, corresponding to different data collection times, the object is reconstructed by the HIO algorithm with additional reality and positivity constraints. The resolution in each reconstructed object is evaluated from the spatial frequency cut off for the corresponding transfer function. Then the counting time scales with resolution as $d^{-3.99}$ in exact agreement with analytic solution. However, this time is found to be higher than predicted, likely because variations in scattered intensities rather than absolute count rate must be reliably detected. That requires collecting the higher number of counts at the maximum scattering angle. The data acquisition time can be reduced by using lower X-ray energy, increasing the number N of simultaneously exposed molecules or addition of convex constraints (such as histogram) to reduce the degree of oversampling required. Supported by ARO and NSF.

[1]See Starodub *et al.*, J. Chem Phys. 123, 244304 (2005)